B2C 商城畅销电子图书的分布特征与在线评论解析*
——以京东为例

雷兵[1,2]　钟镇[1,2]
[1] 河南工业大学管理学院　郑州 450001　[2] 河南工业大学商务智能与知识工程实验室　郑州 450001

摘要：[目的/意义]通过对京东商城电子图书特征的数据分析，发现和识别不同类型畅销电子图书在定价、编辑、出版、销售与用户体验等方面的共性特征与差异性。[方法/过程]借助网络爬虫，采集京东商城 7 个类别销量居前 60 位的电子书的售价、纸版定价、纸版售价、文件格式与大小、评论数量与评论文本等信息，通过文本挖掘、描述性统计等方法，考察畅销电子书的定价策略、销售数量、文本格式与用户体验。[结果/结论]畅销电子书定价普遍较低，不同类别、特征图书定价略有区别，专业类图书定价高于大众类图书；电子书与纸质书的受众群体存在差异，电子书的销量与其纸版的销量并非成正比关系；EPUB 流式文件和 PDF 版式文件占主导地位，两种格式的用户接受度无明显差异；畅销电子书的用户体验主要集中在“满意度”“价格”与“便捷”3 个方面。

关键词：电子图书　畅销书　定价策略　销售数量　文件格式　用户体验　评论挖掘
分类号：G305
DOI：10.13266/j.issn.0252-3116.2018.21.007

chinaXiv:202308.00514v1

1　引言

随着电子文档技术的发展和便携式阅读终端的普及，电子图书（以下简称“电子书”）市场规模不断增大。根据中国新闻出版研究院统计数据，2016 年中国电子书市场收入已达到 52 亿元。然而，与数字出版业的整体规模相比，电子书占比还很小，仅为 0.91%[1]。这说明图书出版行业在数字化融合中还需要加大发展力度。电子书的发展受多种因素的影响，既包括阅读终端、文件格式、定价策略、内容生产、版权保护、市场运营等供应端因素[2]，也包括阅读习惯、技术采纳等读者行为因素[3]。对于图书情报学科而言，电子书读者的使用行为、阅读习惯与主题偏好具有哪些显著特征则是学科关注的重点。

京东（JD.COM）是中国知名 B2C 商城，其图书类目凭借品牌、流量、物流、推广等优势，迅速拓展线上市场。2017 年第三季度，京东图书市场占全国线上图书销售市场份额的 36.2%，位居国内线上图书销售首位[4]。在电子书方面，不同于亚马逊（Amazon）采取以 Kindle 阅读器为中心构建相对独立的电子书生态系统，京东更强调兼容性：一方面以 EPUB 和 PDF 等主流文件格式进行编辑出版，降低了图书出版商制作电子书的门槛和成本；另一方面开发了能安装在不同操作系统（如 Windows、Android 与 iOS 等）的多版本客户端软件以及便携式“水墨屏”电子书阅读器，适应多种阅读环境，降低了读者的购买与使用成本。

本文基于电子书读者采纳行为视角，以京东为例，通过数据采集、统计分析以及评论文本挖掘，对 B2C 商城电子书销售情况展开实证研究，解析当前不同类别畅销电子书在定价策略、销售数量、文本格式与用户体验等方面的共同特征与差异性，以期为电子书阅读行为与动机的理论研究提供经验证据。

2　文献回顾

2.1　电子书定价的相关研究

定价是电子书的一个研究热点，已有研究主要从

* 本文系国家自然基金项目“作者、期刊与数据库错误引文的科学计量学研究：识别方法、产生机理与抑控对策”（项目编号：71603073）和河南省高校社科创新团队支持计划项目“在线知识产品定价机制与付费行为研究”（项目编号：2014-CXTD-06）研究成果之一。
作者简介：雷兵（ORCID：0000-0002-1073-4724），副教授，博士；钟镇（ORCID：0000-0001-6248-2226），副教授，博士，通讯作者，E-mail：zhongzhen@haut.edu.cn。
收稿日期：2018-04-18　修回日期：2018-06-23　本文起止页码：53-61　本文责任编辑：易飞

ChinaXiv合作期刊

定价权、定价影响因素以及定价策略等方面展开。电子书的销售方式主要有买断制、批发制和代理制 3 种[5]。L. Hao 和 M. Fan 研究了批发制和代理制模式下的电子书定价问题,认为采用高价格的代理制降低了电子书的总销售收入[6];Q. Lu 等建立了由作者、出版商、实体书店和电子书店构成的供应链模型[7];B. D. L. Santos 和 M. R. Wildenbeest 运用亚马逊等 B2C 商城的销售数据实证检验了代理制模式对电子书 B2C 市场定价的影响[8]。国内方面,胡兴球和曲文凤认为电子书定价的影响因素主要包括成本、定价权及版权保护等[9],而巫国义则认为还包括利益目标、市场需求、市场竞争等因素[10];此外,王璐璐和徐振云分析了电子书免费的动因,包括供给、需求、品牌等方面[11],董小梅从出版社角度提出了电子书独立定价、订阅定价及拆分定价策略[12];叶兰等将电子书定价分为按集成库销售、按本购买、逐年订购等 7 种模式[13];刘银娣通过比较中美电子书定价制度,提出采用固定价格制度、基于印刷图书价格制订电子书定价方案、基于大数据进一步优化价格等定价策略[14]。

2.2 电子书销量的相关研究

该领域研究主要聚焦于定价、促销以及纸质图书销售等因素对电子书销量的影响。D. Bounie 等运用计量经济学方法,对 2007 年 11 月至 2010 年 7 月间亚马逊网站上纸质版、电子版畅销书进行对比分析,发现电子书与纸质图书的销量并不具有一致性[15];H. Chen 等采用自然实验研究方法,对亚马逊网站 2010 年间纸质版先于电子版发布和纸质版与电子版同时发布的图书进行对比实验,发现推迟电子版发布导致其电子版销售额下降的同时却未带来纸质版图书销售额的增长,并据此认为短期内电子书对纸质图书的冲击不大[16]。在国内电子书零售平台方面,许洁等曾基于对“多看阅读”电子书页面报价信息的统计分析,发现图书制作成本对电子版定价的影响不显著,但打折比例影响电子版销量[17];石玺则运用营销学4Ps 理论,对以淘宝、当当、京东、亚马逊等电子商务平台的电子书营销策略进行对比分析,提出了相应的建议[18]。

2.3 电子书格式的相关研究

电子书的格式纷繁复杂,既有专用格式,又有开放格式,标准很难统一,主要包括 PDF、LIT、AZW、EPUB、TXT、CAJ 等,以 PDF 和 EPUB 应用最为广泛[19]。PDF 属于版式文件,适合图、表、公式等较多的专业图书,但无法自适应不同尺寸的显示屏,尤其是在移动终端上体验不佳;而 EPUB 属于流式文件,可以根据阅读终端显示屏尺寸自动重排,但对图、表、公式的展示有所不足[20]。目前,EPUB3.0 以上版本基于 HTML5 技术,已经能够支持 SVG(可缩放矢量图)、MathML(数学标记语言)等功能,提升了对教育、科技类文档的支持度[21]。

2.4 电子书用户体验的相关研究

电子书用户体验的研究主要集中在阅读终端对用户体验的影响以及购买体验等方面。T. Nakajima 等[22]对 6 种 EPUB 电子书便携式阅读终端进行了测试,覆盖了 iOS、Android、Windows 及 Kindle Linux 操作系统,研究结果显示 iPad 平板电脑的 iBooks3 阅读器效果最佳;Y. Zeng 等[23]基于阅读实验和问卷调查,发现在 PC 机、笔记本电脑等传统大显示屏中版式文件较受欢迎,而在便携式移动终端中流式文件更受用户偏爱。N. Amblee 和 T. Bui 以亚马逊网站低价电子书(Shorts E-Books)为例,指出网络口碑能够传递电子书内容、作者以及同类电子书的声誉,并建议销售商和读者都应将网络口碑作为社交购物体验的首要来源[24]。

综上所述,国内外学者对于电子书的研究内容已十分广泛,深入到定价、销量、格式及用户体验等多个方面。尽管如此,目前仍有一些可以改进的方面:①国外实证研究以亚马逊等主流线上商城数据为主,研究内容较为深入,而国内研究则以定性阐述为主,实证研究明显不足;②在线评论与用户体验之间的关联性分析存在部分缺失,缺乏针对评论文本中关于电子书用户体验方面的特征提取分析。

3 研究设计

京东商城于 2012 年 2 月正式启动电子书刊业务,销售平台与智能手机/PC 阅读客户端软件 LeBook 同步上线。经过多年发展,目前已经成为国内最大的电子书销售平台。京东电子书丰富的品类以及较为广泛的受众群体使其成为研究我国电子书出版与销售现状的绝佳数据来源。

3.1 数据来源与采集方法

本研究的数据获取自京东数字内容频道(http://e.jd.com),数据采集时间为 2018 年 1 月 20 日,采用八爪鱼采集器 7.2.4 版(http://www.bazhuayu.com),通过采集规则配置、云采集等步骤,获取本研究的原始数据,采集过程分成基本信息采集与评论本文采集两个步骤,样本容量见表 1。

第一步采集的是畅销电子书基本信息,包括书名、ISBN、分类、电子版实洋、纸版码洋和实洋、文件格式、

雷兵，钟镇. B2C 商城畅销电子图书的分布特征与在线评论解析——以京东为例[J]. 图书情报工作,2018,62(21):53－61.
ChinaXiv合作期刊

表 1　数据采集数量

类别	大众图书			教育图书	专业图书			合计
	文学	少儿	居家生活	教材考试	社会科学	管理	工业科技	
电子书数量(种)	59	59	60	60	59	60	60	417
采集评论数(条)	5 890	4 168	2 102	3 553	5 855	5 915	2 380	29 863

文件大小、电子版评论数、纸质版评论数 9 个数据项。按照京东数字内容频道分类,每类采集电子书评论数排名前 60 位(京东每页显示 60 个产品)的图书。京东电子书未采用《中国图书馆图书分类法》,而是自主划分为文学、管理、社会科学、工业科技、居家生活、教材考试、少儿 7 类,因此本研究共采集 420 条记录。7 类图书中,文学、居家生活、少儿 3 类属于大众图书,教材考试类属于教育图书,工业科技、管理和社会科学属于专业图书。数据获取后,经过人工审核,发现文学类第 54 条、社会科学类第 48 条以及少儿类第 10 条为广告产品,予以剔除,因此样本实际包含的是 417 本电子书。

第二步采集的是畅销书的用户评论文本。在第一步采集的畅销书中,最高评论数为 4 100,而最低为 10。考虑到研究的实际需求及京东网站"反爬"机制的限制,本文对于评论数不足 100 条的全部予以采集,而对于超过 100 条的评论采集前 100 条。

3.2　数据分析维度与方法

根据本文的研究目标以及采集到的数据记录与属性,从定价、销量、格式以及用户体验 4 个维度分析不同类型畅销电子书的特征分布。以下对定价、销量、格式 3 个维度的分析方法给出说明,用户体验维度的分析方法见 3.3。

(1)定价分析。通过采集数据获取到电子书价格以及对应的纸质书定价和价格,以纸质书定价为基准,通过描述性统计分析,比较不同类别的电子书定价策略。在统计分析时,采用"电子书价格 = 纸质书定价 × 电子书折扣""纸质书价格 = 纸质书定价 × 纸质书折扣"进行描述。

(2)销量分析。通过采集数据分析不同类别、不同价位电子书的销量情况。京东网站上未提供图书的累计销量,但提供了累计评价数。京东的产品评价是针对已购买产品的消费者设置的一项功能,包括评分和文字评论。根据自愿原则,消费者可以对产品进行评论,从而使产品页面生成一条评论文本,累计评论数量加 1;消费者也可以不评论,累计评论数量不增加。本文以累计评价数量近似代表电子书累计销量的做法,理由有 4 点:①当消费者购买商品后,京东网站采取发送提醒信息、奖励"京东豆"(京东网站的虚拟货币)等手段促进购后评价,其评价率较高;②本研究的目标消费群体均为电子书消费者,属于同类评价群体,其购后评价率相近;③J. Chevalier 和 D. Mayzlin 通过对亚马逊网站图书数据的回归分析发现,购后评论数量与图书销量有显著正相关关系[25],国内学者许洁等也采用评论数替代销量的做法[17];④对本文获取的数据分析发现,417 本畅销电子书中,评论数最小值为 10,最大值为 4 100,且最小间距为 10,标准差为410.9,这说明评论数的离散度较大,因此评论数较高者其销量较高的概率很大。

(3)格式分析。京东电子书有 EPUB(流式)与 PDF(版式)两种格式,通过描述性统计,分析不同类别图书两种格式的占比、文件大小以及不同格式的销量情况。

3.3　用户体验分析方法

本文通过对购后评论文本的分析,获取电子书用户体验情况,包括不同类别、价位、格式电子书的用户体验特征分布。京东电子书用户评论文本主要包括针对图书内容与针对用户体验两种类型。本文主要聚焦于电子书介质的用户体验,因此评论挖掘的重点放在与电子书整体满意度相关的特征词分析上。主要工具为 R 语言及相关开源工具包,分为 3 个步骤:

(1)中文分词。对采集到的 29 863 条评论文本,利用开源分词软件包——"结巴中文分词"(jiebaR0.9.1),选取其中的混合模型算法进行分词。为提高准确性,构建了本地词库,加入 5 个专有名词,包括"电子书""京东阅读器"等近义词的合并处理。同时,借助《四川大学机器学习智能实验室停用词库》进行停用词去除,最终得到 16 029 个不同的关键词。

(2)基于词频与评论覆盖率的特征词筛选。首先对 16 029 个关键词进行词频统计,使用工具包 wordcloud 制作词云,并进行词频统计。然后,取频度最高的前 380 个关键词,这些关键词的评论覆盖率均在 0.2% 以上(即,每 500 条评论至少出现 1 次),将其作为特征词的备选集合。最后,人工筛选剔除评论图书内容及无意义的词语,得到与电子书用户体验相关的 33 个关键词,作为用户体验特征词。

chinaXiv:202308.00514v1

ChinaXiv合作期刊

(3)特征词的分类。通过对表 2 中 33 个评论涉及电子书用户体验(Experience)的高频特征词的分析,发现可以将其划分为满意度(正面 + 负面)、价格感知与便捷性体验 3 个单一类别。其中,满意度(Satisfaction)是对购买的电子书的整体评价,包括正面和负面两种情况,如"电子书不错,省了纸质的麻烦""电子版,我喜欢哈哈哈""千万别购买电子书! 很垃圾,还不退!"等。需要说明的是,为了提取与"用户体验"而非"图书内容"相关的满意度特征词,本研究借助表 2 中最左侧一栏的"评价对象",依据两组词汇的"共现"性(两者是否出现在同一分句),来判别满意度关键词是否为用户体验特征词。价格(Price)与便捷(Handle)属于较为直接的用户体验词汇,在识别时不再进行与"评价对象"共现对比。当每条文本中出现价格类特征词时 $Price=1$,否则 $Price=0$;$Handle$ 变量的赋值做同样处理。凡涉及 3 类特征词中任何一类的评论,本研究均将该评论的 $Experience$ 赋值为 1,否则为 0(见公式(1))。

表 2 特征词一览

评价对象词汇	满意度词汇		价格类特征词	便捷类特征词
	正面	负面		
电子版、电子书、电子书、京东阅读、客户端	不错、还行、还好、好好、好用、很棒、很好、太好、满意、可以、值得、喜欢、超值	不好、垃圾	便宜、划算、价格、免费、实惠、物美价廉、优惠、性价比、特价	方便、速度、物流、下载

$$Experience=\begin{cases}1 & (polarity=1)\ or\ (polarity=-1)\ or\ (price=1)\ or\ (Handle=1)\\ 0 & \text{其他}\end{cases} \quad 式(1)$$

为了评估特征分类方法的有效性,本文随机抽取 100 条评论文本,依据"真实语义"统计了 3 类特征词的正确率、召回率及 F 值(计算公式:正确率 * 召回率 *2/正确率 + 召回率),分别达到 89.83%、91.38% 与 90.60%,证明了该方法是有效的。

4 结果与分析

4.1 电子书价格分布特征

定价策略是影响电子书销量的重要因素,而价格与折扣则是表征定价策略的两项重要指标。从表 3 可以看到,417 本畅销电子书平均价格为 7.17 元,相对于纸版书的平均折扣为 2 折。当前国内 B2C 商城的纸版图书折扣一般为 6–8 折,如此低的折扣(售价低)显示电子书还处于推广期,低价策略是其主要推广方式之一。7 个类别中,管理类、工业科技类、社会科学类等专业类图书的总体折扣较高(售价相对较高),均在 2.4折以上;教材考试类、文学类、居家生活类、少儿类折扣较低(售价相对较低),均未超过 1.6 折。这种折扣上的差异性表明,专业图书消费者对于电子书的价格接受程度要高于大众图书、教育图书。

表 3 畅销电子书的价格与折扣 (单位:元)

类别	文学	少儿	居家生活	教材考试	社会科学	管理	工业科技	全体
电子版平均售价	6.71	3.13	4.17	4.83	9.02	15.29	7.00	7.17
纸版平均定价	44.12	20.15	27.59	35.09	37.82	53.93	26.64	35.16
电子版总体折扣	0.15	0.16	0.15	0.14	0.24	0.28	0.26	0.20

在具体折扣的选择上,如表 4 所示,50% 以上电子书的折扣在 2 折及以下;还有约 20% 左右的电子书定价保持在 2–4 折区间,剩下不到 20% 的图书在 4–6 折以及 6 折以上区间。换言之,80% 以上的京东畅销电子书都以不到其"码洋"一半的价格进行销售。

京东畅销电子书如此低的折扣率,缘于其中包含了大量的定价在 0–1 元之间"超低价格"电子书。这类图书合计 104 本,占总数的 24.94%,主要由"版权过期"图书构成,以少儿图书为主,也有类似《三国演义》与《基督山伯爵》这样的文学名著。不难理解,纸版书生产成本主要由版权费、制作费和材料费构成,电子书没有材料费,如果版权免费,其边际成本接近于 0,因而可以超低价销售。从表 4 我们同样能看到,虽然数

表 4 京东畅销电子书折扣分布

类别 \ 折扣	≤2 折	2–4 折	4–6 折	>6 折
文学	69.49%	27.12%	1.69%	1.69%
少儿	71.19%	25.42%	1.69%	1.69%
居家生活	68.33%	30.00%	1.67%	0.00%
教材考试	78.33%	18.33%	0.00%	3.33%
社会科学	57.63%	28.81%	6.78%	6.78%
管理	56.67%	25.00%	15.00%	3.33%
工业科技	63.33%	18.33%	16.67%	1.67%

chinaXiv:202308.00514v1

ChinaXiv合作期刊

量十分有限，但仍有一些电子书折扣在6折以上，价格与纸版图书的“实际售价”（实洋）相差无几。高折扣（即高价）电子书主要分为两种情况：一种是热销纸版书的电子版，如《长尾理论》、《腾讯传（1998－2016）中国互联网公司进化论》等，市场需求促使定价较高；另一类是读者群体为小众的专业图书，纸版书已停止再版但仍有一定的市场需求，如浙江大学出版社于2009年出版的《基于Xilinx FPGA的OFDM通信系统基带设计》、人民邮电出版社于2011年出版的《新手配色宝典4：室内设计篇》等。“1元封顶”及6折以上图书的数量分布情况如图1所示：

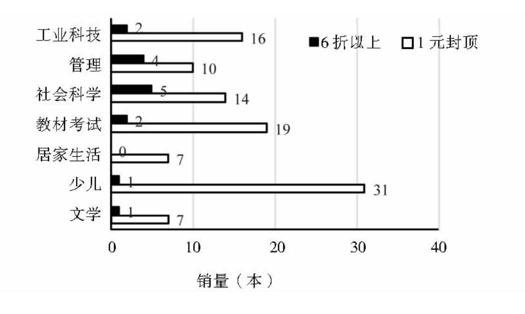


图1 “1元封顶”及6折以上电子书的分布

4.2 基于评论数的电子书销量分布特征

借助评论数，可以了解京东不同类别畅销电子书的销量特征。对比表5畅销书电子版与纸版的评论数均值，发现京东畅销电子书的销量呈现出以下3个特征：

（1）电子/纸质评论数均值的比值高达2.26%，说明现阶段纸版书仍是图书消费的主流形式。

（2）电子书的销量与纸质书的销量没有显著的正相关关系。评论数均值显示，电子书各分类中，管理类与社会科学类销量最高；而对应的纸版图书中，销量最高的则是文学、社会科学与少儿类。

（3）电子书销量对图书总销量的影响在不同类别之间差异明显。工业科技评论数均值为112，但其电子版/纸版评论数比值高达27.11%；电子版销量较高的社会科学类，电子版/纸版的比值却只有2.00%，对图书总销量的影响反而不如销量较少的一些图书。

针对电子书销量的第二和第三个特征，在不考虑图书内容主题的影响因素之下，一种解释是：当前电子书还处于推广期，市场份额较小，读者对电子书的接受程度影响电子书的销量。

表5 京东畅销电子书与对应纸质图书的评论数均值

类别	文学	少儿	居家生活	教材考试	社会科学	管理	工业科技	全体
电子版评论数均值	303	125	54	180	407	458	112	234
纸版评论数均值	21 353	16 320	689	2 108	20 366	11 501	411	10 328
电子版/纸版（%）	1.42%	0.77%	7.88%	8.55%	2.00%	3.99%	27.11%	2.26%

值得注意的是，在比较畅销电子书与其纸质版的过程中，我们还发现两类十分有趣的图书：“绝版”的纸质图书（即平台无纸版销售）与超过纸版销量的电子书（见图2）：①无纸版销售的电子书共计74本，均价9.37元，高于畅销书电子版7.17元的平均水平，占总体的比例为17.75%。这类图书的主要特征为出版年份较早且未再版，而较低的市场销售量也不足以驱动出版社再版印刷（如湖北人民出版社的《希腊神话故事新编（1979年）》等）。其中，教材考试类较多，文学类最少。②超过纸版销量的电子书共计124本，均价为4.47元，低于畅销电子书的整体均价，占总体的比例为29.74%。这类图书的纸版虽然还在销售，但销量十分有限，因而出版商与在线平台愿意将其电子版低价出售，一方面引导读者接受电子书的消费形式，另一方面提升滞销纸版图书的销量。在这些低价畅销书中，工业科技类最多，有30本，占该类别的一半；社会科学最少，仅10本。不难理解，对于“绝版”与“低销量”的纸版书而言，它们的“电子化”再销售实际拓展了消费市场，而不是通常所认知的“取代”。

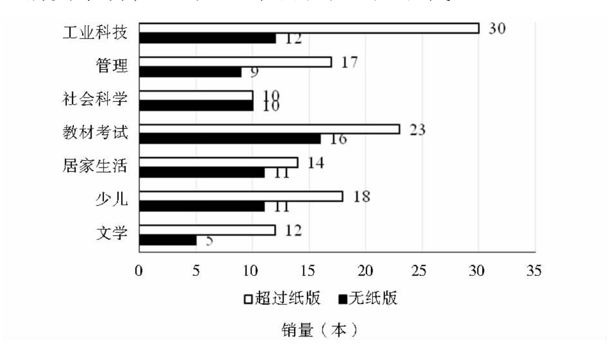


图2 超过纸版销量与无纸版销售的京东畅销电子书数量分布

尽管低价策略在理论上可以提高销量，但电子书的推广还是应当立足于满足消费者的内容需求与良好体验，而不能简单地依靠低价渗透方式。实际上，通过比较不同价格区间评论数均值，可发现京东畅销电子书的定价与销量并非一条斜率为负的曲线。如图3所示，将畅销电子书的售价分为（0,1]、（1,2]、（2,5]、

chinaXiv:202308.00514v1

ChinaXiv合作期刊

(5,10]、(10,20]、(20,40]和 >40(单位:元)7 个区间:1 - 2 元之间的图书评论数均值最低(117 条/本),向左定价低至 1 元以内,评论数均值增长了 1.74 倍;但向右定价增长的同时,评论数均值不但没有降低,反而呈线性增长。进一步考察评论数超过 1 000 的电子书(合计 24 本),可发现其主要由社会科学类、管理类及工业类图书构成。可以想象,该类图书消费者对电子版的接受程度较高,且购买图书的主要目的是学习知识,对于定价或折扣比例的敏感度相对较低。例如,管理类图书《奇点系列(套装 6 册)》纸版定价 200 元,电子版售价 147.1 元,折扣 7.16 折,然而评论数高达 1 700个,位居 417 本电子书评论数的第四位。

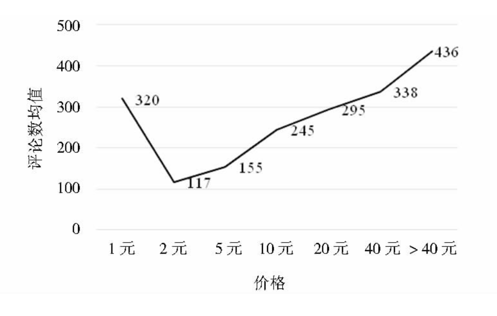


图 3 不同价位电子书的评论数均值

4.3 电子书文件格式分布特征

417 本畅销电子书中,EPUB 格式图书为 239 本,占 57.31%;PDF 格式图书为 178 本,占 42.69%。在文件大小方面,417 本电子书均值为 21.97MB,EPUB 格式图书均值为 6.98MB,而 PDF 格式图书为 42.08MB,是 EPUB 格式的 6 倍。这主要缘于两个格式的内容存储方式上的差异,EPUB 的文字存储趋近于"文本",而 PDF 的文字存储趋近于"图片"。

从表 6 可以看到,社会科学类、文学类及管理类图书以 EPUB 格式为主,其内容多为文字,采用 EPUB 格式恰好满足了读者自适应各种屏幕大小阅读终端的需求;而工业科技类、教材考试类及居家生活类图书中 PDF 格式占绝大多数,大都含有一定数量的图片、数学公式,而这些内容在 EPUB 格式中体验并不好,故出版机构选择了 PDF 版式结构。此外,也存在有部分以文字为主的以 PDF 版式结构出版的电子书,如部分少儿图书与文学名著。它们所用纸版的出版日期较早,京东采用低价策略销售(均价为 0.79 元),通过简单扫描纸版书制作电子版,没有进行较为复杂的 EPUB 格式转化。

表 6 京东畅销电子书不同文件格式的图书数量、评论数与容量均值

类别		文学	少儿	居家生活	教材考试	社会科学	管理	工业科技	全体
PDF	图书数量(种)	3	18	46	51	2	6	52	178
	评论数均值(条)	163.33	125	40.65	88.63	400	316.67	122.88	79.78
	文件大小均值(MB)	143.10	69.73	60.25	25.03	129.90	32.68	25.05	6.98
EPUB	图书数量(种)	56	41	14	9	57	54	8	239
	评论数均值(条)	310.71	124.88	99.29	700	406.84	474.07	37.5	348.62
	文件大小均值(MB)	5.76	3.64	24.92	1.58	5.04	6.70	23.13	42.08

需要说明的是,本研究分别对比了 EPUB 与 PDF 两种格式图书的定价与评论数均值,发现差异并不显著,说明文件格式对图书定价与销量的影响十分有限。

4.4 电子书用户体验的分布特征与影响因素

基于本文 3.2 所述方法,对 29 863 条评论文本进行了分词,并基于评论中的高频词汇绘制了京东畅销电子书评论的词云图(见图 4),文字大小代表关键词出现的频度。词频统计结果显示,排名前 10 位的词语依次为"不错、很好、电子书、可以、好好、值得、喜欢、非常、京东、方便",均为正面观点词汇,表明消费者对畅销图书的满意度较高,这与 417 本图书评价分均值为 4.85(满分 5 分)的结果是一致的。

在这些评论中,与用户体验相关评论共有 5 380 条,占到总评论数的 18.02%,而剩下的超过 80% 的评论或没有实际意义(如默认好评等),或只有图书内容的评价,难以测度用户体验。本文将具有"用户体验性

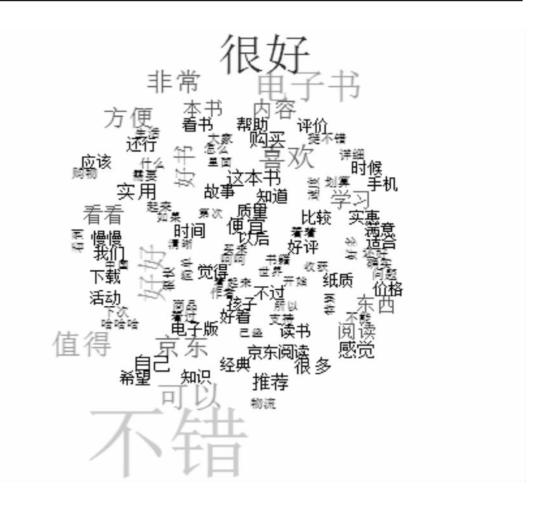


图 4 在线评论文本词云

chinaXiv:202308.00514v1

ChinaXiv合作期刊

特征”的评论称为“体验性评论”,体验性评论占总评论数的比例称为“体验性评价率”,表 7 列出了 3 类体验性评论数量与 7 类电子书的体验性评价率。从类别上看,工业科技类的体验性评论比例最低,仅为 13.28%,这可能与该类图书消费者对电子书接受程度较高有关,导致其评论更多围绕图书内容展开。3 类体验性评论中,满意度评论共计 1 616 条,评价率为 5.41%,包括 1 548 条正面评论与 68 条负面评论,较高的正面满意比例说明电子书体验总体良好。而在为数不多的“不满意”评论中,焦点主要集中阅读器的限制与 PDF 文件相对较差的体验特征,如“只支持京东阅读器”“PDF 阅读体验不好”等。但限用专门阅读器是电子书零售市场常见形态(如亚马逊),而选择 PDF 还是 EPUB 格式则取决于出版商提供的原始文件格式以及图书内容。价格类与便捷类的总体评价率分别为 8.28% 和 8.15%,高于平均满意度,其原因在于不少评论仅仅做了类似“电子书不错”这样的评论,而未表明满意来自价格还是便捷性。

表 7　京东畅销电子书体验评论分布数量与体验性评价率

类别	满意度				价格		便捷		体验	
	正面		负面							
文学	344	5.84%	9	0.15%	527	8.95%	410	6.96%	1057	17.95%
少儿	233	5.59%	8	0.19%	284	6.81%	366	8.78%	765	18.35%
居家生活	78	3.71%	5	0.24%	155	7.37%	168	7.99%	333	15.84%
教材考试	174	4.90%	18	0.51%	284	7.99%	391	11.00%	705	19.84%
社会科学	336	5.74%	14	0.24%	509	8.69%	501	8.56%	1069	18.26%
管理	311	5.26%	11	0.19%	583	9.86%	432	7.30%	1135	19.19%
工业科技	72	3.03%	3	0.13%	132	5.55%	167	7.02%	316	13.28%
合计	1548	5.18%	68	0.23%	2474	8.28%	2435	8.15%	5380	18.02%

统计结果表明,产品定价、文本格式与文件大小是影响体验性评价的主要因素。在价格方面,如表 8 所示,10 元以下低价图书对于电子书的体验评价率不足 20%,而 10 元以上图书的评价率在 44% 以上。对这种显著性差异的一种可能解释是:当电子书的价格较低时,消费者对电子书的期望度不高,因此相关评论数量较少,这从部分涉及体验评论的文本也可看出,如“五毛钱很划算,支持京东阅读,书也不错”“电子书,内容还可以,随便看看”等;而当电子书的价格较高时,消费者较关注电子版介质的用户体验,将电子版与纸版书作对比,如“电子书比纸质书便于携带,不错的”“还好,有配图,电子版价格实际不太便宜”等。

表 8　不同价位图书评论文本语义特征分布

价格区间(元)	满意度		价格	便捷	体验
	正面	负面			
≤1	5.50%	0.21%	9.28%	8.67%	19.70%
(1,5]	4.49%	0.35%	7.54%	7.79%	16.89%
(5,10]	5.55%	0.06%	8.61%	8.10%	17.94%
(10,40]	14.67%	0.38%	20.09%	20.41%	44.45%
>40	14.07%	0.00%	21.10%	19.27%	45.57%

在文件格式方面,EPUB 格式与 PDF 格式的电子书体验评价率分别为 18.00%、18.06%,无明显差异。但在满意度方面,EPUB 格式的正面评价比 PDF 格式好一些,高出 1.07%,这与 EPUB 格式对不同尺寸显示屏的适应度高有关。文件大小方面,50MB 以内的图书与 50MB 以上的图书在便捷方面的评价略有差异,如某消费者购买文件大小为 185.78MB 后的评论为“下载太耗流量了”,但整体上针对文件大小的评论文本不多,这与当前中国 WIFI 和 4G 普及、网速较快关系密切,文件大小对用户体验的影响不大。相关统计数据分别如表 9 与表 10 所示:

表 9　不同文件格式图书评论文本语义特征分布

文件格式	满意度		价格	便捷	体验
	正面	负面			
EPUB 格式	5.50%	0.19%	8.52%	7.63%	18.00%
PDF 格式	4.43%	0.33%	7.73%	9.38%	18.06%

表 10　不同文件大小图书评论文本语义特征分布

文件大小	满意度		价格	便捷	体验
	正面	负面			
≤10M	5.23%	0.18%	8.70%	7.66%	5.23%
≤50M	4.86%	0.32%	6.85%	8.43%	4.86%
≤100M	5.22%	0.28%	8.25%	10.92%	5.22%
>100M	6.08%	0.58%	7.60%	10.41%	6.08%

5　结论与讨论

本文基于京东网站数据,从定价策略、销售数量、文本格式与用户体验 4 个方面,考察了不同类别畅销电子书的共性特征与差异性,主要研究结论如下:

chinaXiv:201308.00514v1

ChinaXiv合作期刊

(1)畅销电子书定价普遍较低,不同类别、特征图书定价略有区别。当前电子书市场处于推广期,相对于纸版书而言,读者的接受程度还不高,低价策略能够促进电子书的销售,如无版权费图书以“一元封顶”的超低价拓展电子书市场。管理、工业科技、社会科学等专业类图书读者群体的电子版接受程度高于其他图书读者群体,定价略高于文学、教材考试等大众图书和教育图书。

(2)畅销电子书的销量与其纸版的销量并非成正比关系。一般认为,电子书的推广将蚕食纸版图书市场。然而,该观点并未得到本文的实证支持。417 本图书中,接近一半的电子书销量并不依赖于纸版书,这些电子书的销量要么高于纸版书,要么是其纸版书没有在京东网站上销售。该实证结论说明电子书与纸质书的受众群体存在差异,推广电子书能够拓展图书市场空间,提升出版商和 B2C 商城的销售收入。

(3)畅销电子书中,EPUB 流式文件和 PDF 版式文件占主导地位,两种格式的用户接受度、满意度无明显差异。出版商对于电子书格式的选取主要取决于图书内容,以文字为主的图书多采用 EPUB 格式,而图片、公式较多的图书以 PDF 格式为主。畅销电子书中还存在一种现象,出版日期较早的图书,其纸质版已无销售,出版商将其制作成 PDF 纸质扫描版,B2C 商城将其以超低价销售,作为推广电子书消费市场的一种促销手段。

(4)畅销电子书的用户体验特征主要聚焦于满意度、价格及便捷 3 方面。读者对于电子书的整体满意度较高,超过 95%;读者非常关注电子书的价格和便捷度,“便宜”和“方便”是读者对电子书最直观的认识。不同类别图书消费群对于电子书用户体验的关注度略有区别,其中工业科技类图书读者关注度最低,这可能与该类图书群体对电子书接受程度较高有关;高价位电子书的用户体验关注度明显高于低价位电子书,一种可能的原因是当电子书价位较高时,更易与纸质版对比从而提高电子版的接受程度。

限于技术手段,本文研究存在一定的局限性:一是选取的数据仅为畅销电子书,没有包含纸质版畅销书,其样本类型不够丰富;二是以电子书评论数近似代表累计销量,这导致统计分析的准确度降低;三是以在线评论作为电子书用户体验的观测点,不能全面反映消费者的真实想法。针对这些问题,将在未来的研究中做进一步的完善。

chinaXiv:202308.00514v1

参考文献:

[1] 中国数字出版产业年度报告课题组.“十三五”开局之年的中国数字出版——2016 - 2017 中国数字出版产业年度报告主报告(摘要)[J]. 出版发行研究, 2017(7): 5 - 10.
[2] 张美娟, 潘涵, 黄龙舟. 近年来我国电子书产业研究述评[J]. 出版科学, 2013, 21(6): 48 - 50.
[3] 赵文军, 陈焕之, 蒋伟进. 近 10 年来国内外数字阅读研究综述[J]. 图书情报工作, 2017, 61(18): 128 - 136.
[4] 易观. 中国 B2C 市场季度监测报告 2017 年第 3 季度 [EB/OL]. [2018 - 03 - 10] https://www.analysys.cn/analysis/22/detail/1000995/.
[5] 郑丽芬. 从定价博弈看电子书产业发展[J]. 中国出版, 2014(4): 11 - 14.
[6] HAO L, FAN M. An analysis of pricing models in the electronic book market[J]. MIS quarterly, 2014, 38(4): 1017 - 1032.
[7] LU Q, SHI V, HUANG J. Who benefit from agency model: a strategic analysis of pricing models in distribution channels of physical books and e - books[J]. European journal of operational research, 2018, 264(3): 1074 - 1091.
[8] SANTOS B D L, WILDENBEEST M R. E - book pricing and vertical restraints[J]. Quantitative marketing & economics, 2017, 15(2): 85 - 122.
[9] 胡兴球, 曲文凤. 电子书定价的价格歧视策略探讨[J]. 科技与出版, 2013, 32(8): 14 - 18.
[10] 巫国义. 电子书定价策略评析[J]. 科技与出版, 2014(10): 95 - 98.
[11] 王璐璐, 徐振云. 电子书免费动因分析[J]. 编辑之友, 2015(7): 36 - 39.
[12] 董小梅. 出版社电子书定价策略创新探究[J]. 科技与出版, 2015(3): 56 - 58.
[13] 叶兰, 孙坦, 黄国彬, 等. 电子图书定价模式研究[J]. 图书情报工作, 2007, 51 (12): 81 - 83.
[14] 刘银娣. 中美电子书价格现状与定价制度比较研究[J]. 编辑之友, 2015 (4): 105 - 108.
[15] BOUNIE D, EANG B, SIRBU M, et al. Superstars and outsiders in online markets: an empirical analysis of electronic books[J]. Electronic commerce research and applications, 2013, 12(1): 52 - 59.
[16] CHEN H, HU Y J, SMITH M D. The impact of ebook distribution on print sales: analysis of a natural experiment [EB/OL]. [2017 - 12 - 14]. http://dx.doi.org/10.2139/ssrn.1966115.
[17] 许洁, 丁乐华, 彭威. 基于“多看阅读”的电子书定价规律数据挖掘分析[J]. 信息资源管理学报, 2017 (1): 89 - 96.
[18] 石玺. 基于电子商务平台的电子书出版分析——以淘宝阅读、当当数字馆、京东电子书刊、kindle 电子书店为例[J]. 新闻界, 2015 (2): 56 - 59.
[19] 郑滢瑜, 丛挺, 缪婕. 电子书格式标准研究[J]. 出版科学, 2010, 18(4): 90 - 93.
[20] 陈敏. 专业类电子书的特点及其消费者行为分析——基于化

ChinaXiv合作期刊

学工业出版社的实证研究[J]. 科技与出版，2012（5）：64－67.

[21] IDPF. EPUB 3.1[EB/OL].[2018－03－10]. http://idpf.org/epub/31/.

[22] NAKAJIMA T, SHINOHARA S, TAMURA Y. Typical functions of e－textbook, implementation, and compatibility verification with use of epub3 materials[J]. Procedia computer science, 2013, 22: 1344－1353.

[23] ZENG Y, BAI X, Xu J, et al. The influence of e－book format and reading device on users' reading experience: a case study of graduate students[J]. Publishing research quarterly, 2016, 32(4): 319－330.

[24] AMBLEE N, BUI T. Harnessing the influence of social proof in online shopping: the effect of electronic word of mouth on sales of digital microproducts[J]. International journal of electronic commerce, 2011, 16(2): 91－114.

[25] CHEVALIER J A, MAYZLIN D. The effect of word of mouth on sales: online book reviews[J]. Journal of marketing research, 2006, 43(3): 345－354.

作者贡献说明：

雷兵：负责论文框架确定、资料搜集与整理、数据分析、初稿撰写；

钟镇：负责提出选题、设计研究方案、撰写论文、修改论文并定稿。

Research on Distribution Characteristics of Best Selling eBooks on B2C Store and Their Customer Reviews of Experience: A Case Study of JD.com

Lei Bing[1,2] Zhong Zhen[1,2]

[1] School of Management, Henan University of Technology, Zhengzhou 450001

[2] Business Intelligence & Knowledge Engineering Lab, Henan University of Technology, Zhengzhou 450001

Abstract: [**Purpose/significance**] Pricing, edition, publishing, selling and customer experiences of best selling ebooks in JD.com were analyzed for the aim to discover and identify general characteristics and differences of best sellers in ebooks. [**Method/process**] The Web crawler was employed to collect information of firt 60 best selling ebooks in 7 different categories. The information involved include price of ebooks, price of paperbook, cost of paper book, file format and size, number and text of comments. Some methods, such as descriptive statistics and text mining, were used to study pricing strategies, sales quantity, text format and user experience of best selling ebooks. [**Result/conclusion**] Many ebooks are sold in lower price than paperbooks, but there are a little difference between different categories, which the professional ebooks' price are higher than popular ebooks. There are some differences between ebooks and paperbooks on audiences, the positive relationship is not significant between ebooks and paperbooks. The EPUB and PDF are the main text format, they have no obvious difference in customer acceptance. The consumers' comments, which contain user experience, focus on satisfaction, price and handling.

Keywords: ebooks best selling books pricing strategies sales quantity text format user experience review mining

chinaXiv:202308.00514v1

下　期　要　目

□ 城市图书馆年度报告制度研究　（何盼盼　陈雅）
□ 新一代学术交流生态系统的研究与构建——以北京大学为例　（崔海媛　罗鹏程　聂华等）
□ 中美公共图书馆未成年人服务实践比较研究——基于网站内容的分析　（雷雪　李丹）
□ 学术图书馆研究数据服务的多元路径　（涂志芳　徐慧芳）
□ 国际图联国际宣传项目的分析与推广　（何韵　程焕文）
□ 美国高校图书馆版权信息服务的调研与思考　（孙涵涵　郝群　张立彬）